

WHAT IS CLAIMED IS:

1. A method for routing communications in a storage system, comprising:

5 assigning a preference to each of a plurality of communication paths between a source node and a destination node, wherein each preference indicates a relative measure of how often one of the communication paths should be selected for sending one of a plurality of communications, and wherein each preference indicates that a corresponding communication path should
10 be chosen at least once when sending the communications;

receiving one of the communications to be sent from the source node to the destination node;

15 selecting a communication path from the plurality of communication paths between the source node and the destination node;

20 sending the communication on the selected communication path from the source node to the destination node; and

repeating said receiving, said selecting, and said sending for a plurality of communications to be sent from the source node to the destination node, wherein said selecting is performed so that a more preferred path is selected more often than a less preferred path;

25 wherein each of the communication paths between the source node and the destination node is independent of the other communication paths.

30 2. The method as recited in claim 1, wherein the plurality of communications comprises communications to one or more disk drives.

3. The method as recited in claim 1, wherein said assigning comprises assigning static preference values to each of the communication paths.

5 4. The method as recited in claim 3, wherein said assigning static preference values comprises assigning a higher preference to ones of said communication paths located in a region of the storage system assigned to the source node and assigning a lower preference to ones of said communication paths that connect through a region of the storage system assigned to another source node.

10

5. The method as recited in claim 3, wherein said assigning static preference values comprises assigning static preference values according to predicted traffic patterns, wherein said static preference values are assigned to reduce congestion.

15 6. The method as recited in claim 3, wherein said assigning static preference values comprises assigning static preference values according to predicted traffic patterns, wherein said static preference values are assigned to reduce a probability of deadlock conditions occurring.

20 7. The method as recited in claim 1, further comprising:

maintaining a routing table of the communication paths, wherein the routing table comprises a plurality of entries, wherein each entry indicates one of the communication paths; and

25

wherein said selecting comprises selecting a next entry from the routing table.

8. The method as recited in claim 7, wherein a number of times that one of the communication paths is selected relative to the other communication paths corresponds to a number of entries for that communication path in the routing table.

30

9. The method as recited in claim 8, wherein the number of entries for that communication path in the routing table corresponds to a preference assigned to that communication path.

5

10. The method as recited in claim 1, wherein each preference comprises a ranking of one of the communication paths.

11. The method as recited in claim 1, wherein each preference comprises a percentage
10 of the plurality of communications that should be sent using one of the communication paths.

12. The method as recited in claim 1, further comprising:
15 detecting an error on one of the communication paths during said sending of one of the communications; and

resending on a different one of the communication paths the one of the communications that was being sent when the error was detected.

20

13. The method as recited in claim 1, wherein said selecting the communication path comprises calculating the communication path.

14. The method as recited in claim 13, wherein said calculating further comprises
25 calculating the communication path from a pair of coordinates identifying the source node and the destination node.

15. A storage system, comprising:

a plurality of nodes interconnected by an interconnection fabric, and wherein a portion of the nodes are coupled to one or more storage devices;

a source configured to send a plurality of communications to a destination node,
5 wherein the source comprises a source node and a source device configured to communicate with the source node;

wherein the source is further configured to send the communications from the source node to the destination node by selecting one of a plurality of
10 communication paths between the source node and the destination node according to a preference assigned to that communication path, wherein the source is configured to repeatedly select communication paths until all of the communications are sent, and wherein the source is configured to select a more preferred communication path more frequently than a less
15 preferred communication path;

wherein the source is further configured to select each of the plurality of communication paths at least once and thus to send at least one of the communications on each of the communication paths; and

20 wherein each of the communication paths between the source node and the destination node is independent of the other communication paths.

16. The storage system of claim 15, wherein the plurality of communications
25 comprises communications to one or more disk drives.

17. The storage system of claim 15, wherein the source is further configured to maintain a routing table of the plurality of communication paths, wherein the routing table comprises a plurality of entries, and wherein each entry indicates one of the
30 communication paths.

18. The storage system of claim 17, wherein the source is further configured to select a communication path by selecting a next entry from the routing table.

5 19. The storage system of claim 18, wherein the source is further configured to select one of the communication paths a number of times relative to the other communication paths, wherein the number of times corresponds to a number of entries for that communication path in the routing table.

10 20. The storage system of claim 19, wherein the number of entries for that communication path in the routing table corresponds to a preference assigned to that communication path.

15 21. The storage system of claim 15, wherein the source is further configured to receive an indication that an error occurred on one of the communication paths used to send one of the communications and to resend the communication that was being sent when the error occurred on a different one of the communication paths.

20 22. The storage system of claim 15, wherein the source is further configured to select the communication path by calculating the communication path.

23. The storage system of claim 22, wherein the source is further configured to calculate the communication path from a pair of coordinates identifying the source node and the destination node.

25

24. A device, comprising:

an interface configured to communicate with a source node in an interconnection fabric coupling together a plurality of nodes and providing a plurality of

independent communication paths between the source node and a destination node; and

5 a controller configured to communicate a plurality of communications to the source node to be sent on the independent communication paths to one or more disk drives coupled to the destination node;

10 wherein the controller is further configured to communicate the communications to the source node by selecting one of the communication paths according to a preference assigned to that communication path, encoding the selected path in a communication, and communicating the communication to the source node, and wherein the controller is configured to repeatedly select communication paths, encode communication paths in communications, and communicate communications until all of the communications have
15 been communicated;

wherein the controller is configured to select a more preferred communication path more frequently than a less preferred path;

20 wherein the controller is further configured to select each of the independent communication paths at least once and thus to encode each of the communication paths in at least one of the communications.

25 25. The device of claim 24, wherein the controller is further configured to maintain a routing table of the plurality of communication paths, wherein the routing table comprises a plurality of entries, and wherein each entry indicates one of the communication paths.

26. The device of claim 25, wherein the controller is further configured to select a communication path by selecting a next entry from the routing table.

27. The device of claim 26, wherein the controller is further configured to select one of the communication paths a number of times relative to the other communication paths, wherein the number of times corresponds to a number of entries for that communication path in the routing table.

5

28. The device of claim 27, wherein the number of entries for that communication path in the routing table corresponds to a preference assigned to that communication path.

29. The device of claim 24, wherein the controller is further configured to receive an indication that an error occurred on one of the communication paths used to send one of the communications, to encode a different one of the communication paths in the communication that was being sent when the error occurred, and to recommunicate that communication to the source node.

30. The device of claim 24, wherein the controller is further configured to select the communication path by calculating the communication path.

31. The device of claim 30, wherein the controller is further configured to calculate the communication path from a pair of coordinates identifying the source node and the destination node.

32. The device as recited in claim 24, wherein the controller comprises a RAID controller.

33. The device as recited in claim 24, wherein the controller comprises a host interface configured to communicate with a host computer.

34. The device as recited in claim 24, wherein the controller comprises a disk drive controller.

30

35. A node, comprising:

a routing unit;

5 a plurality of input ports; and

a plurality of output ports;

10 wherein the node is configured to be connected to an interconnection fabric coupling together a plurality of nodes and comprising a plurality of independent communication paths between the node and a destination node;

15 wherein the routing unit is configured to send a plurality of communications on the independent communication paths to one or more disk drives coupled to the destination node;

20 wherein the routing unit is further configured to send the communications to the one or more disk drives coupled to the destination node by selecting one of the communication paths according to a preference assigned to that communication path and sending a communication on the selected communication path, wherein the routing unit is configured to repeatedly select communication paths and send communications until all of the communications are sent; and

25 wherein the routing unit is configured to select a more preferred communication path more frequently than a less preferred path;

wherein the routing unit is further configured to select each of the independent communication paths at least once and thus to send at least one communication on each of the communication paths.

5 36. The node of claim 35, wherein the routing unit is further configured to maintain a routing table of the plurality of independent communication paths, wherein the routing table comprises a plurality of entries, and wherein each entry indicates one of the communication paths.

10 37. The node of claim 35, wherein the routing unit is further configured to select a communication path by selecting a next entry from the routing table.

38. The node of claim 37, wherein the routing unit is further configured to select one of the communication paths a number of times relative to the other communication paths,
15 wherein the number of times corresponds to a number of entries for that communication path in the routing table.

39. The node of claim 38, wherein the number of entries for that communication path in the routing table corresponds to a preference assigned to that communication path.

20 40. The node of claim 35, wherein the routing unit is further configured to receive an indication that an error occurred on one of the communication paths used to send one of the communications and to resend the communication that was being sent when the error occurred on a different one of the communication paths.

25 41. The node of claim 35, wherein the routing unit is further configured to select the communication path by calculating the communication path.

42. The node of claim 35, wherein the routing unit is further configured to calculate the communication path from a pair of coordinates identifying the source node and the destination node.

5 43. The node as recited in claim 35, wherein one of the input ports is configured to communicate with a RAID controller.

44. The node as recited in claim 35, wherein one of the output ports is configured to communicate with a disk drive.

10

44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1